IBM/Coursera Capstone

For the Data Science Professional Certificate

Introduction

- The Premise:
  - We are hired to find the client a new spot to open a Deli in the city of Toronto
  - Client is very good at what he does and is confident that he can out-compete light competition
- Our Strategy:
  - Find areas with good population density to make up the clientele
  - Make sure the competition is light so that our client can take the market by storm

Initial Data

- Information on which neighborhoods existed and what their coordinates were was collected.
- This data was merged into a single data frame so that we could cluster the neighborhoods by GPS coordinates.

```
[4]: geodata = pd.read_csv("Geospatial_Coordinates.csv")
     geodata.rename(columns={"Postal Code":"PostalCode"},inplace=True)
     geodata.head()
```

[4]:

	PostalCode	Latitude	Longitude
0	M1B	43.806686	-79.194353
1	M1C	43.784535	-79.160497
2	M1E	43.763573	-79.188711
3	M1G	43.770992	-79.216917
4	M1H	43.773136	-79.239476

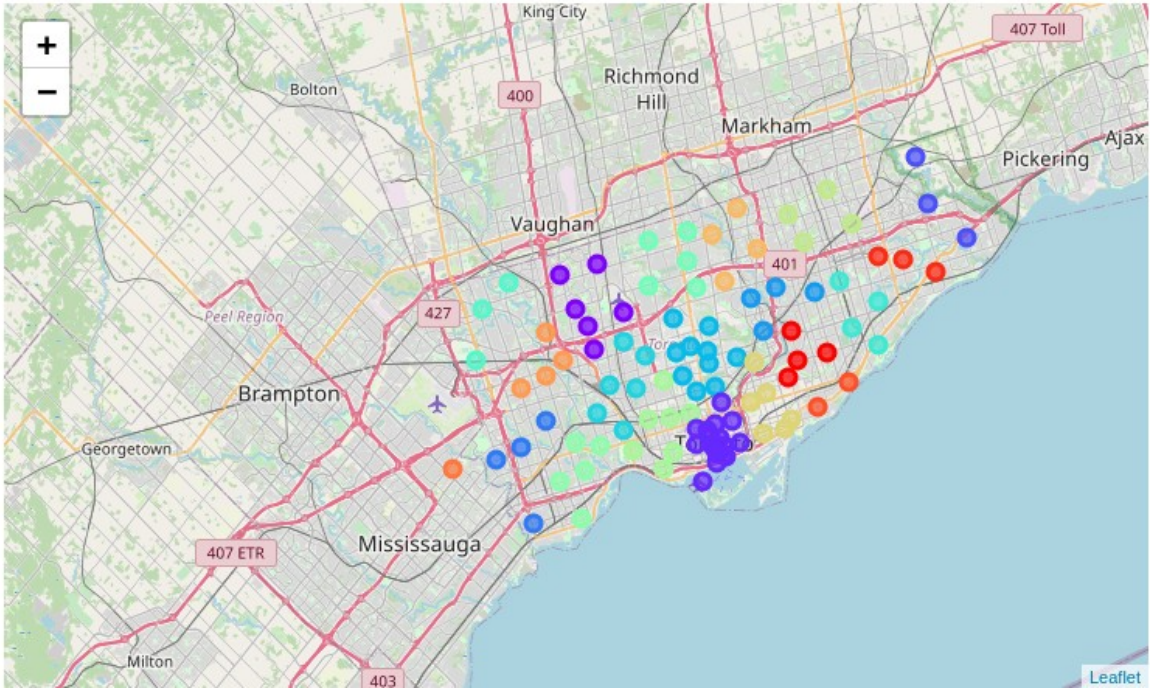
```
[5]: df = df.merge(geodata,on="PostalCode")
     df.head()
```

[5]:

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
3	M6A	North York	Lawrence Manor, Lawrence Heights	43.718518	-79.464763
4	M7A	Queen's Park	Ontario Provincial Government	43.662301	-79.389494

Clustering

- K-Means Clustering Algorithm was used
- Number of clusters, k, was chosen visually using plotted points on the map
- K = 20 was chosen as it resulted in the most clusters without singleton neighborhoods making up a cluster



Aggregation of Data on Clusters

[59]:

	ClusterNumber	Latitude	Longitude	CompetitorCount	AverageRating
0	0	43.709684	-79.307119	22	6.900000
1	1	43.741392	-79.494290	11	0.000000
2	2	43.653767	-79.381828	30	6.640000
3	3	43.809115	-79.186829	1	0.000000
4	4	43.641182	-79.551913	13	8.400000
5	5	43.743784	-79.329654	26	7.133333
6	6	43.708745	-79.396811	30	0.000000
7	7	43.690505	-79.465231	21	0.000000
8	8	43.736597	-79.253571	7	0.000000
9	9	43.734156	-79.582818	7	0.000000
10	10	43.769799	-79.420311	19	0.000000
11	11	43.635194	-79.502445	16	0.000000
12	12	43.662682	-79.428994	30	0.000000
13	13	43.797654	-79.292324	16	0.000000
14	14	43.676923	-79.336286	28	0.000000
15	15	43.781679	-79.367674	15	0.000000
16	16	43.704217	-79.534349	11	0.000000
17	17	43.636966	-79.615819	16	0.000000
18	18	43.684507	-79.278940	14	0.000000
19	19	43.769234	-79.215035	4	0.000000

Clusters were studied based on number of prospective competitors in the area and the average rating of competitors

[31]:

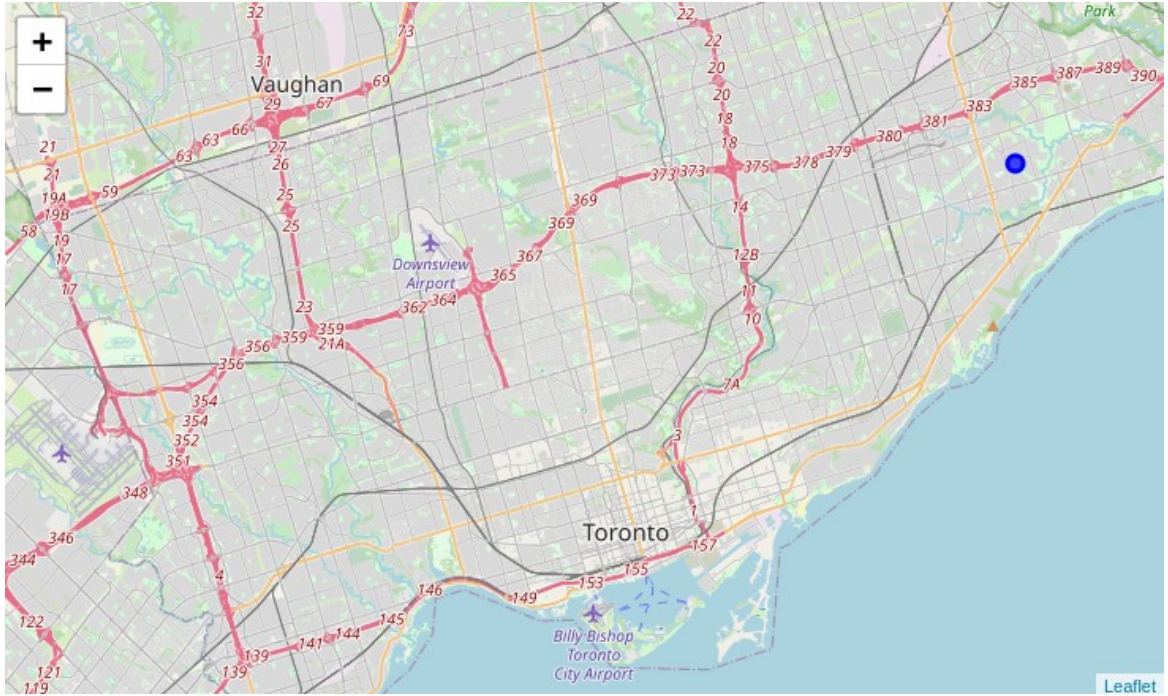
```
search_query = 'Deli'
radius = 3000

DFs = []

for coordinates in k_means.cluster_centers_:
    url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&oa
    result = requests.get(url).json()

    try:
        DFs.append(pd.json_normalize(result['response']['venues']))
    except:
        DFs.append(None)
```

[39]:

```
for each in DFs:
    display(each.head())
    print(each.shape)
```

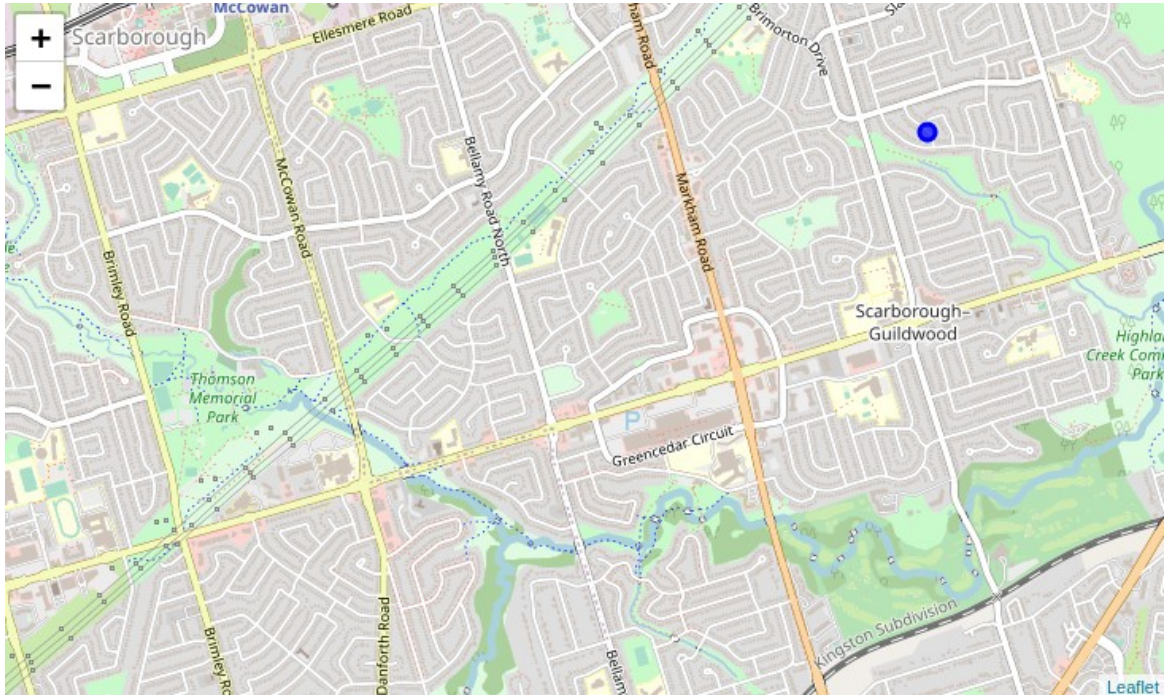
	id	name	categories	referralId	hasPerk	location.address	location.cro
0	4bc5dfbf4a9aa593a00b077b	Dr. Deli & The Salad Queen	[{'id': '4bf58dd8d48988d1c5941735', 'name': 'S...	v-1623927873	False	1881 Steeles Ave W	D
1	4dbf46c41e72dd48b1ef56ef	Oak Park Deli	[{'id': '4bf58dd8d48988d1c4941735', 'name': 'R...	v-1623927873	False	NaN	
2	5954162804d1ae42efb9b367	Druxy's Famous Deli	[{'id': '56aa371be4b08b9a8d573550', 'name': 'F...	v-1623927873	False	1200 Eglinton Avenue East	
3	4f5a4771e4b0357e5d6b1667	Jody's Deli	[{'id': '4bf58dd8d48988d146941735', 'name': 'D...	v-1623927873	False	777 Warden Avenue	

Final Location Chosen



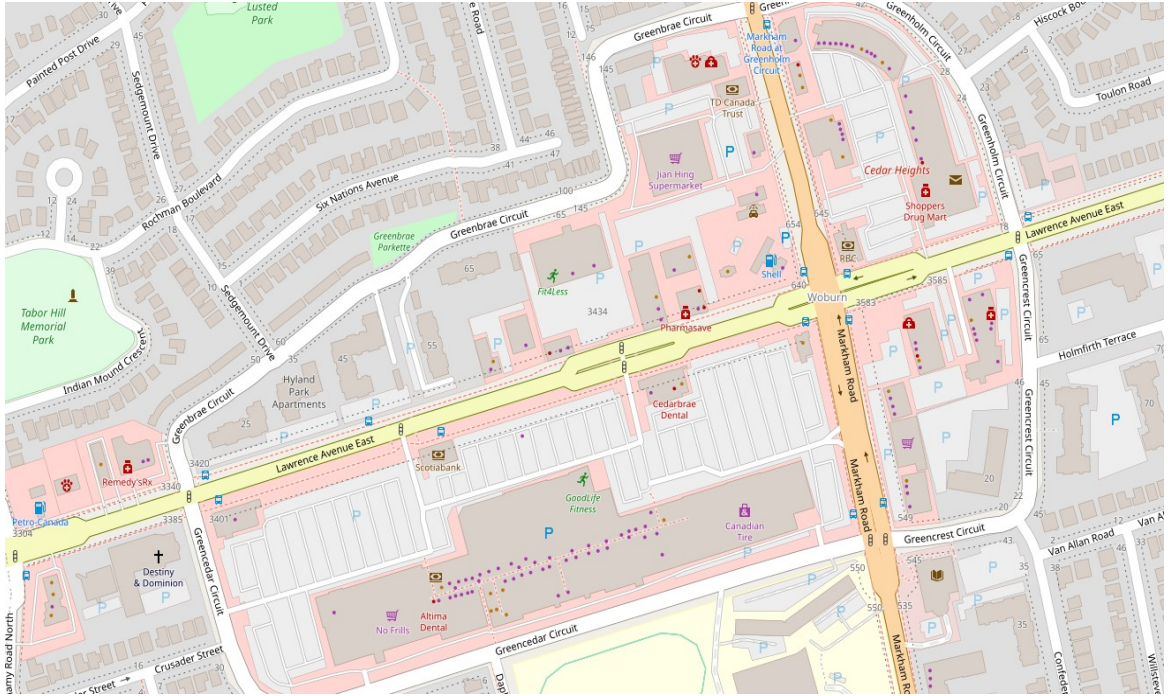
Location on the eastern edge of Toronto was chosen for its apparent lack of competition and lower to nonexistent ratings. Zooming in will reveal more opportunity.

Final Location Chosen



As you can see there is a substantial commerce area surrounded by schools and neighborhoods here.

Final Location Chosen



Zooming in further, we can see that it is quite alive. We have a winner for our location!